Perspective taking and decision-making in educational game play: A mixed-methods study

Lacey J. Hilliard, Mary H. Buckingham, G. John Geldhof, Patricia Gansert, Caroline Stack, Erin S. Gelgoot, Marina U. Bers & Richard M. Lerner


To link to this article: https://doi.org/10.1080/10888691.2016.1204918

Published online: 26 Jul 2016.
Perspective taking and decision-making in educational game play: A mixed-methods study

Lacey J. Hilliard, Mary H. Buckingham, G. John Geldhof, Patricia Gansert, Caroline Stack, Erin S. Gelgoot, Marina U. Bers, and Richard M. Lerner

Tufts University; Oregon State University; Boston University; University of California, Berkeley

ABSTRACT
Video games have the potential to be contexts for moral learning. We investigated whether Quandary, a video game designed to promote ethical thinking and moral considerations for decision-making, would help promote positive skills such as perspective taking and empathy in adolescents. We examined the effect of playing Quandary on 131 middle school students on self-reported measures of moral thinking via mixed-method randomized control trials. In addition, we conducted qualitative analyses of one-on-one participant interviews and short-answer responses to capture experiences and reflections from playing Quandary, as well as the depth in which students across conditions responded to the interview questions. We found that short-term quantitative indicators did not show change across conditions; however, qualitative analyses revealed thematic responses that are consistent with the core components of the Quandary game, and that students in the Quandary condition showed a greater depth of response to interview questions. This work is a first step in exploring the potential for virtual game play on children’s social, emotional, and cognitive development.

Within the last 50 years, video games have developed from a niche hobby into a common and popular form of entertainment. Estimates are that 80% to 90% of children and adolescents play video games, with an average play time of 13.2 hours per week (Gentile et al., 2009). Video games are now the fastest-growing form of entertainment among adolescents, and researchers are beginning to consider ways in which video game play may influence development. In addition, games are becoming increasingly complex, involving more emotionally laden experiences, meaningful narratives, and multifaceted elements than earlier decades focused on puzzles and basic challenges (e.g., Oliver et al., 2015).

Despite this trend toward more complex narratives and serious content, previous research on video games has frequently drawn from a deficit-based perspective. Researchers have asked how video games may lead to negative outcomes such as aggression and addiction, but have paid little attention to the potential positive effects of playing video games. Some work suggests that video games may be useful contexts for learning (e.g., Bers, 2012; Bers & Kazakoff, 2013; Prensky, 2001) and, although still regarded by many as tangential to traditional teaching and learning, gaming during classroom free time is becoming more prevalent (Bers & Kazakoff, 2013; de Freitas & Griffiths, 2008; Prensky, 2001).

The trend toward increasing use of games and simulations for teaching has important implications for understanding how informal and formal learning can support and reinforce one another in order to accelerate learning, support higher-order cognitive development, and strengthen motivation in learning (Bers & Kazakoff, 2013; Delanghe, 2001; Green & Bavelier, 2007). Oliver et al. (2015) describe a recent complexity evolution that modern games have gone through; that is, games are becoming more complex, serious, and lengthy, often involving moral stories and implications. There is a need for baseline research into how games and simulations are currently being used for learning, however. The goal of the present study was to expand this research by considering the potential role of educational video games in promoting facets of moral development.

Quandary is a particular example of an educational online game designed to promote moral decision-making skills by placing youth in the role of a leader who must make difficult decisions to care for a community. We will first discuss video games as a context for learning from a Relational Developmental Systems (RDS) metatheoretical perspective. We then present our research on youth engagement with Quandary and how playing Quandary might influence empathy, perspective taking, and helping behavior. Finally, we...
address less successful aspects of the game and possible improvements.

**Theoretical perspectives**

Models of human development derived from RDS emphasize mutually influential person-context relations (depicted as individual ↔ context relations), or “developmental regulations,” as the focus of development (e.g., Lerner, 2002; Lerner, Lerner, Bowers, & Geldhof, 2015). The double-sided arrow highlights the simultaneous impact that individuals have on their context and that the context has on individuals (Brandstädter, 1998; Overton, 2015). When there is a good fit between the needs, motivations, and abilities of the person and the context in which the person is situated, these developmental regulations have the potential to be mutually beneficial for both person and context. In that case, they are termed “adaptive developmental regulations” (Brandstädter, 1998).

Video games can be seen as virtual contexts that mutually interact with players. The game presents players with a scenario in which some set of actions are possible. The player then chooses which actions to take and the game scenario changes depending on the player’s actions (Isbister & Schaffer, 2008). This exchange may create a pleasurable experience that continually changes as the game adjusts based on the player’s actions and the player makes decisions based on the scenarios the game provides (Isbister & Schaffer, 2008; Przybylski, Rigby, & Ryan, 2010). This mutual influence allows games to provide content, levels of challenge, and options that are tailored to suit the needs of individual players.

Our focus here is on how video games provide a context for moral and ethical thinking. Kohlberg (1987) describes justice and fairness as key principles of moral reasoning. Damon (2008) argues that educational and scholarly efforts on moral development should focus on contexts relevant to youth and to expand what educators and researchers define as moral contexts (for instance, not confining moral discussions to specifically designed education curricula). Furthermore, he encourages educators and scholars to consider the myriad of components related to the morality of justice, such as considering other’s fortunes and misfortunes and actively offering support (i.e., to make the link between care and responsibility). In this article, we draw on these ideas and the tenets of RDS to explore the role of perspective-taking, decision-making, and consideration of others in a video game context.

**Video games as contexts for learning**

Many of the characteristics that make a commercial video game successful are also key components of learning (Bers & Kazakoff, 2013; Gee, 2005). Progressing through a video game typically involves facing increasing levels of challenge. Video games can be designed to keep players motivated by presenting obstacles that may be too difficult to overcome without focused effort and scaffolding, but not so difficult that the player cannot make any progress at all. Furthermore, if an individual is trained to do a task or solve a problem in only a single way, then he or she may struggle to transfer skills to a new context or problem. Because of this design, research has shown that skill training in video games can transfer to real-world contexts (e.g., Anderson, Bavelier, & Green, 2010; Anguera et al., 2013).

These two principles also contribute to a third aspect of games as a learning context: gaining expertise in skills and being challenged to combine them in novel ways in order to overcome new obstacles (Lerner & Callina, 2014). Learners become experts when they practice skills sufficiently to master them, and then use those skills in contexts that require adjusting and reconsidering those skills (Halverson, 2005). In many video games, a level will require players to make use of a set of skills many times, and success will require mastering those skills. Players are then confronted with a final situation, in which they must use the skills they have mastered in new ways to overcome a new and more difficult obstacle.

These are few examples of how the design elements that make a video game engaging and successful are also important aspects of successful learning. Research has found that video games can be useful contexts for learning both cognitive and task-specific skills such as attention, memory, self-control, and performance in surgical simulations (e.g., Anderson & Kirkorian, 2015; Anguera et al., 2013; Basak, Boot, Voss, & Kramer, 2008; Brown et al., 1997; Rosser et al., 2007). In fact, Gee (2005) has argued that not only are video games conducive to learning, but in fact well-designed video games inherently involve learning because of the navigation and problem-solving skills involved in playing them. However, the potential of video games as contexts for moral learning remains rather unexplored, despite the growing call for moral or character education curricula both in school-based and out-of-school time programs (Berkowitz, 2012). Some researchers have found that playing video games with prosocial gameplay elements increases prosocial behavior in subsequent tasks and reduces aggressive cognitions (Gentile et al., 2009; Greitemeyer & Osswald, 2009; Sestir & Bartholow, 2010).

Furthermore, video games can have moral meaning and significance to players (Hartmann, Toz, & Brandon, 2010; Hartmann & Vorderer, 2010). In some cases, players may feel guilty when engaging in unjustified
violence in video games (e.g., against harmless characters), and these feelings are especially strong for players with high empathy (Hartmann et al., 2010; Hartmann & Vorderer, 2010). Furthermore, many games are designed with these moral components in mind (Boyan, Grizzard, & Bowman, 2015; Joeckel, Bowman, & Dogruel, 2012). In such games, players may or may not be sensitive to explicit moral issues, yet increased moral salience—awareness of moral issues—has been linked to decreased moral violations (Joeckel et al., 2012). Therefore, explicit introductions and discussions of game themes may be important for games designed to foster moral growth. These studies further suggest that players treat video games as meaningful virtual worlds, feeling empathy for the game characters and thinking about decisions as they would in real life. It may therefore be possible to design video games that promote moral learning.

The present study

Video games have potential as a context for learning, but until recently, the existing research has focused primarily on academic learning (e.g., de Freitas & Oliver, 2006). Here, we extended prior research by considering the potential benefits of Quandary, an educational video game designed to promote moral learning. We addressed two main research objectives. First, we examined how youth engaged with Quandary. In order for a video game to be a positive context for moral learning, youth must find the game engaging and enjoyable (Mitchell & Savill-Smith, 2004). The game must also successfully communicate the themes and ideas it is intended to teach (Mitchell & Savill-Smith, 2004; Olson, 2010). From one-on-one interviews, we explored youth reactions to Quandary and themes that emerged from their discussion of what they learned by playing this game. Second, we tested whether playing Quandary led to improvements in empathy, perspective taking, and helping behavior, key components of moral behavior. We hypothesized that children who played Quandary would show greater improvements in scores on our measures of these constructs compared to those who played a comparison game; further, we predicted that these improvements would be greater for youth in a condition that included reflection through group discussions. In short, we used both qualitative and quantitative analyses in the present research. We should note that we employed qualitative analyses to assess whether it was triangulate findings across quantitative and qualitative analyses and, as well, to ascertain if qualitative analyses would illuminate the thinking of participants in ways survey-based quantitative scores could not.

Method

Participants

Students in Grades 6, 7, and 8 were recruited from three northeastern U.S. schools to participate in this study. Researchers sought schools who were racially, ethnically, and socioeconomically diverse. Within each school, researchers presented the study to principals, who helped to identify teachers who would be interested in having their students participate. Once schools and teachers agreed to participate, researchers randomly assigned participants to one of three groups: (described in the following section; N[Quandary] = 63; N [Quandary + Facilitation] = 49; N[Control] = 51), and ensured that the three groups did not significantly differ in regard to ethnicity ($\chi^2 (df = 12) = 13.04, p = .37$), gender ($\chi^2 (df = 2) = 0.66, p = .72$), age ($F(2,159) = 0.63, p = .53$), or time spent playing video games each day ($F(2,160) = 2.08, p = .13$). Parental consent forms that contained a brief overview of the study were sent home with the students. Families were given two weeks to have their children return signed parental consent forms to their school directors. Prior to the pre-test questionnaire, students provided assent to participate. We recruited an initial pool of 167 participants at pretest, 163 of whom provided useable data, and 131 of whom provided data at both time points. Our sample was ethnically diverse (8.70% African American, 42.86% Caucasian, 7.45% East Asian, 16.15% Hispanic, 6.83% South Asian, and 20.71% Multiethnic or Other), and represented both genders about equally (51.53% female). The mean age of the participants was 12.79 years ($SD = 1.56$).

Procedures

Our design entailed mixed-method randomized control trials that included quantitative data collected at baseline and after all game play sessions were complete, as well as qualitative data gathered through short answer responses and one-on-one participant interviews. Participants were randomly assigned to one of three conditions and played one of two computer games for a period of four to five weeks. At two schools (School 1 and 2), assignments were made at the group level such that all youth within a classroom were in the same condition. At the third school (School 3), data collection was conducted after school due to the preference of school administrators. In this case, all participants were randomly assigned to one of the three conditions and

---

1The appropriateness of the use of this statistic is questionable because the expected value of 52% of our cells was less than five. We therefore report this statistic for descriptive purposes only.
were matched across gender and grade level. Game play location varied across condition: in School 1 and 2, participants took the surveys and played the game on researcher-provided Chromebooks in their classrooms. In School 3, surveys and game play occurred across multiple locations (including a computer lab, library, and classroom). Researchers facilitated each session; however, teachers were always present. The researchers who facilitated the discussion sessions (i.e., those in the Quandary + Facilitation group) completed a comprehensive training led by a member of the research team to ensure consistency across researchers. Following the post-test assessment, a random selection of students was selected for one-on-one interviews with trained researchers. Interviewers asked the participant to report on their experience playing either Quandary or the control game, and, as well, to respond to questions regarding comprehension of one of the survey scales (on empathic concern).

Quandary

Quandary is a free online game that uses engaging storylines and characters to challenge youth to make difficult ethical decisions that require players to consider the perspectives of others (Learning Games Network, 2012). The Quandary website states that the central game objective is to “strengthen the moral compass of players, by developing the skills that help them recognize ethical issues and deal with ethical situations in their own lives. These skills include: critical thinking, perspective taking, and decision-making (www.quandarygame.org).” In Quandary, the player takes the role of the captain of a space colony on the planet Braxos and, as captain, he or she must investigate and ultimately decide upon solutions to problems that face the colony. These problems are the titular quandaries: there are no clear solutions, but the player’s decisions will have serious consequences for everyone in the colony. In each episode, four possible solutions are available, each with two endings: one in which the solution is successful and improves colony morale and another in which the solution is implemented imperfectly and although the colony solves the problem, morale is damaged. Which ending is presented depends on the player’s performance in the third stage of the game: if the player fails to successfully identify two responses in favor and two responses opposed to the chosen solution, the morale-damaging ending will be shown. Each episode can be played in about twenty minutes.

Each of the three currently available episodes presents a different problem. In the first episode, Little Lost Sheep, the colony’s flock of sheep is being attacked by native predators. The colonists must decide how to protect their livestock. In the second episode, Water Wars, the public colony well has become polluted and the colonists must decide how to restore their water supply. In the third episode, Fashion Faction, the colony tailor has begun making special alterations to his friends’ colony uniforms, causing other colonists to feel excluded.

Game conditions

Students were randomly assigned to one of three conditions: Quandary, Quandary + Facilitation, and Control. In the first condition (Quandary), students played all three episodes of Quandary. Each session lasted approximately 20 min and sessions were separated by approximately one-week intervals; all sessions followed the same sequence of game episodes.

As with the Quandary condition, students assigned to the Quandary + Facilitation condition completed all three episodes of the Quandary video game in sessions that were separated by one week. In addition to playing the game, however, students assigned to the Quandary + Facilitation condition participated in group-level discussions after completing each episode. These discussions were led by members of the research team and followed lesson plans designed to accompany the episode of Quandary that students completed that day. The lesson plans and discussion questions were standard across conditions, though invariably the discussion varied by student responses. Examples of discussion questions included: “What information did you weigh to make your decision? What was the outcome of your solution? How did others in the community respond to your decision? What could have been better? Can you come up with any other solutions to the dilemma?” Discussion times ranged from 5 to 15 min, depending on time constraints in the school schedule.

The third group did not play any episode of Quandary. To ensure that the experiences of the control group closely matched those of the Quandary group, these participants played a different computer game matched to Quandary on design features and engagement level but that lacked a moral component. The game that fit our criteria, Mayan Mysteries, produced by Dig-It! Games (2012), is a puzzle-based game focused on geography and ancient artifacts. The graphics and comic-like appearance are similar to Quandary, and both games include multiple sections with different gameplay. To identify a comparable game, we reviewed games to match style (e.g., comic book appearance, text with read-aloud options) and conducted initial pilot testing to confirm that engagement, play time, and
interest in playing Mayan Mysteries were comparable to playing Quandary. Post-test measures of game liking and interest to play again showed no difference between playing Mayan Mysteries vs. Quandary.

Research sessions
Participation in this study took place across five sessions, each separated by a one-week interval. The first and last sessions only entailed data collection, and participants played either Quandary or Mayan Mysteries during the remaining three sessions. We provide more detail about these sessions in the following sections.

Session 1. Youth completed pre-test measures via Qualtrics, an online questionnaire program. The questionnaire contained demographic items, including race and gender, and quantitative measures, described in detail below.

Sessions 2–4. Youth played either Quandary or Mayan Mysteries for a minimum of twenty minutes on individual Google Chromebooks. In addition to playing Quandary, youth in the Quandary + Facilitation condition participated in group-level discussions following the game play. During these discussions, researchers asked students to respond to questions about the content of the episode, the decisions that they selected throughout the experience, and what possible alternative decisions that they could have made. Students reflected on the character perspectives and the outcome of the characters and community.

Session 5, Part 1. To assess changes in youth characteristics, youth completed the same measures during Session 5 as were administered in Session 1. The post-test questionnaire also contained questions regarding experience playing either Quandary or Mayan Mysteries. In addition, participants watched a brief animated comic and were asked to reflect on its content (described as follows).

Session 5, Part 2. Immediately after completing the questionnaire during Part 1 of Session 5, a subset of participants engaged in one-on-one interviews with researchers about their game play experience, how they think about themselves, and a cognitive interview about how they understood questionnaire items. Participants who indicated they were interested in participating in the interview were randomly selected such that an equal number of participants in each condition were interviewed. These interviews took place in the student’s classroom or in an adjacent classroom. Individual interviews were audiotaped and transcribed for further content analyses.

Qualitative data sources
Qualitative data collection included one-on-one interviews and short answer questions about the game experience.

Interviews
Participants were randomly selected across conditions to participate in interviews (59 interviews in total; see Table 1 for gender breakdown by condition). Interviews were semi-structured, giving interviewers the freedom to inductively explore participants’ responses. As such, although interviewers followed a similar protocol, follow-up questions during each interview varied slightly from participant to participant. These interviews took place during the final study session, after the administration of the post-study questionnaires. Interviews contained two parts: game-related questions and empathy-related questions. During the game-related questions, researchers asked students to discuss if and what they liked about the game (either Quandary or Mayan Mysteries), what they did not like about the game, what they remembered most, if and what they learned from the game, and if they would play the game in their free time. In the empathy-related part of the interview, we read students items from the questionnaire that were from the Davis (1980) empathic concern subscale and asked how they interpreted the response scale, if they recalled ever experiencing feelings and situations described in the items, and how they remembered learning about the empathy-related content. Students responded to three items randomly selected from the full 7-item scale.

Short answer questions
As part of our final survey, we asked students to tell us their favorite episode of Quandary, why they liked it, and their general feelings about Quandary. Students typed their responses to an open-ended prompt in Qualtrics during their final questionnaire session.

Quantitative measures
We examined the effect of playing Quandary on participants’ self-reported levels of positive youth development (PYD), active and engaged citizenship, empathic concern, perspective taking, interpersonal generosity, commitment to moral action, and moral reasoning.

Table 1. Interview participant breakdown across conditions.

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th>Girls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quandary only</td>
<td>9</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>Quandary + Facilitation</td>
<td>10</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Control</td>
<td>10</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>30</td>
<td>59</td>
</tr>
</tbody>
</table>
We detail the measures used to operationalize these constructs below, but note here that we omitted all reverse-coded items from our analyses due to participants’ difficulty in distinguishing these items from non-reverse-coded items. We report Cronbach’s (1951) coefficient α as an estimate of all measures’ internal consistency, with estimates presented as: α at pre-test/α at post-test.

**Positive youth development**

We operationalized positive youth development (PYD) using a modified version of the very short measure of the Five Cs of PYD discussed by Geldhof et al. (2014). The Five Cs model defines PYD as comprised of Five Cs (Competence, Confidence, Character, Caring, and Connection), and our modified measure indexed the Five Cs using 17 items administered on a 5-point Likert-type scale. Competence was indicated by three items that measured academic, social, and physical competence, respectively (α = .56/.52). Social confidence was indicated by three items representing self-worth, positive identity, and physical appearance, respectively (α = .75/.80). Character was indicated by four items representing social conscience, values diversity, conduct behavior, and personal values, respectively (α = .67/.70). Caring was represented by three items that indexed participants’ empathic responding (α = .81/.82). Connection was represented by four items that indicated participants’ connection to their families, peers, schools, and neighborhoods, respectively (α = .77/.81). We also computed an average of all indicators, as an index of each participant’s overall level of PYD (α = .89/.91).

**Empathic concern**

We measured empathic concern using the four nonreverse-coded empathic concern items from the Davis (1980) Individual Reactivity Index (α = .74/.74). An example item is, “I would describe myself as a pretty soft-hearted person.”

**Perspective taking**

We measured perspective taking using the five nonreverse-coded perspective taking items from the Davis (1980) Individual Reactivity Index (α = .83/.78). An example item is, “I believe there are two sides to every question and try to look at them both.”

**Interpersonal generosity**

We measured generosity using the 10-item Interpersonal Generosity Scale (Smith & Hill, 2009; α = .91/.86). An example item is, “I am known by family and friends as someone who makes time to pay attention to others’ problems.”

**Commitment to moral action**

We measured participants’ commitment to moral action using the 8-item Adapted Good-Self Assessment (Barriga, Morrison, Liau, & Gibbs, 2001; α = .91/.92). This measure asks participants how important it is for them to display several positive qualities (e.g., “Considerate or Courteous” and “Fair or Just”).

**Moral reasoning**

We assessed participants’ moral reasoning using the Moral Values Internalization Questionnaire (Hardy, Padilla-Walker, & Carlo, 2008). This scale presents participants with six scenarios (cheating on a test, lying, being mean, making fun of someone, stealing, and not paying someone back after borrowing money from them), and asks why participants would not engage in these behaviors. For each scenario the reasons listed include the importance of the reasons for lying that are: (a) purely selfish, (b) to keep one’s image, (c) following a rule, and (d) identity protecting. We summed all responses related to each reason as measures of our participants’ moral reasoning (α s for selfish: .86/.92; image: .91/.92; rule: .91/.92; identity: .92/.93).

**Cyberball**

The Cyberball task was originally created by Williams, Cheung, and Choi (2000) for use in studying ostracism (see Scheithauer, Tiger, & Miller, 2013, for a review of the use of this task in developmental research). In the present study, we used a variation of Cyberball as a measure of helping behavior. Cyberball takes the form of a simple game in which the participant plays catch with three other characters. The participant views three players, labeled Player A, Player B, and Player C, tossing a ball to each other. Player A and Player B will toss the ball to Player C only once, and then never again. The purpose of this was for the participant to observe what he or she believes is an online game in which one player is being excluded. The participant’s computer then “connects to the game” and his or her character appears on screen. Player A and Player B will now toss the ball among themselves and the participant. During this section, Player A and Player B will throw the ball to the participant 50% of the time and to each other 50% of the time, while Player C will throw the ball to the participant 50% of the time and Player A and Player B 25% of the time each. Throwing targets were chosen via random number generation. After the participant has tossed the ball twenty times, the game ends and the participant is prompted to save a data file, which

---

3Post-survey interviews showed that students had difficulties with interpreting the reverse-coded items in the scale.
contains a list of the participant’s ball-tossing targets. Researchers then debriefed participants and explained that they were not playing with actual players.

**Qualitative plan of analyses**

Interviews were transcribed and content-coded by members of the research team. The interviews were coded inductively to generate information about participant experience playing either game. Specifically, the content and analyses of our interviews focused, first, on aspects of game engagement (i.e., what components resonated well with youth, what did youth like about playing the game, and what did they learn from playing the game) and, second, on how they discussed issues of empathy and perspective taking.

The analysis of the empathy-related questions involved coding the depth of response to the questions. We used a global coding rating based on depth of response and compared levels of depth across Quandary and Control conditions. Specifically, we used a coding system to explore how enriched the students’ responses were across the empathy-related interview questions. Answers were coded as either 1: None/Very Little Depth (i.e., youth gave brief/one-word answers to the questions in the section; provided little to no explanation or detail in responses); 2: Some Depth (i.e., youth gave brief answers to the questions in the section; provided some explanation or detail in responses); and 3: High Depth (i.e., youth gave thorough answers to the questions in the section; provided explanations, details, and examples in responses).

**Quantitative plan of analyses**

Our quantitative analyses can be divided into two parts. The first set of analyses compared our measures across the three groups at baseline to ensure that any differences observed during post-test were not due to group differences that existed prior to playing Quandary or the control game. For these analyses we tested five models: one that compared Global PYD and the Five Cs across groups; one that compared AEC, civic participation, and civic duty across groups; a third model (henceforth called the Prosocial model) that compared measures of empathy, perspective taking, generosity, and commitment to moral action across groups; a fourth, moral reasoning model that compared the four measures of moral reasoning across groups; and a final fifth model that examined the Cyberball task. The first four models consisted of a MANOVA with omnibus Type-I error rate set to .15. We chose this Type-I error rate such that follow-up $T^2$ tests comparing each group would have a Type-I error rate of .05 after implementing a Bonferroni correction. The fifth model that examined the Cyberball task only was an ANOVA with Type-I error rate set to .05.

The next step of analysis involved testing the effect on condition from pre- to post-test; post-test assessments occurred no more than one-week after the final game-play session. The second set of models consisted of doubly multivariate MANOVAs and a repeated-measures ANOVA that examined changes in our outcomes between pre-test and post-test, aggregating outcome variables in five models as was done in the pre-test analyses. We set omnibus Type-I error rates to the same levels as in our pretest analyses and only included participants who provided data available at both time points in these analyses.

Missing data can negatively impact the results of MANOVA models, and the literature strongly recommends the use of advanced analytic techniques when the percentage of missing data is not low (e.g., maximum likelihood estimation, multiple imputation; Enders, 2010). When missingness is low (e.g., < 5%), however, traditional methods of handling missing data (e.g., list-wise deletion, mean substitution) are often considered reasonable (Graham, Cumsille, & Elek-Fisk, 2003). In the present data, total missingness at pre-test was 2.36% (range: 0.00%, 5.52%), with only six variables displaying univariate missingness greater than 5%. For our longitudinal analyses, total missingness was 1.82%, with no variables displaying univariate missingness greater than 5% (range: 0.00%, 4.58%). Given the relatively small amount of missingness, we therefore took a variant of the mean substitution approach. For each scale we computed a participant’s scale score as the mean of all non-missing data points. All analyses then imposed list-wise deletion for participants missing any of the relevant scale scores.

**Results**

In this study, we investigated the use of a moral learning game, Quandary. Our goals were to understand how youth engage with Quandary; what youth learned from playing the game; whether scores on a number of measures related to moral thinking and behavior improved after playing Quandary; and how design elements of the game did and did not successfully engage youth. See Table 2 for a breakdown of coding categories of interview responses in Quandary conditions.

**Engagement with Quandary**

As part of our final survey, we asked students to tell us their favorite episode of Quandary, why they liked it, and their general feelings about Quandary. Students
Table 2. Coding categories of interview responses in Quandary conditions.

<table>
<thead>
<tr>
<th>Category</th>
<th># Instances coded across interviews</th>
<th>Percentage across interviews *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game-related components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision-making</td>
<td>26</td>
<td>33</td>
</tr>
<tr>
<td>Leadership</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Morality-related components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Importance of community</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Consideration of others</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>Perspective-taking</td>
<td>20</td>
<td>25</td>
</tr>
</tbody>
</table>

*Percentage of interviews covering at least one category: 92%.

found Quandary fairly challenging (mean difficulty rating = 3.51, out of a possible score of 5; from not at all challenging to extremely challenging), and were somewhat likely to play the game again for fun or recommend it to friends (mean = 3.08 and 3.11, respectively, out of a possible score of 5). Nearly half of the students listed Episode 3, Fashion Faction, as their favorite episode (47 listed Episode 3 vs. 23 for Episode 1 and 36 for Episode 2). Many of the reasons given involved a personal interest in the topic or being able to relate to the story. For example, some reasons were: “because my school is getting uniforms and I think my school should listen to what we think about them,” “it was the one that I felt connected the most to me—no one wants to be left out,” “Because I like how the citizens are all about trying something new and want to be treated equally and it’s about clothes,” and “Because I’m a girl and I love fashion. And I could relate to that episode.” Similarly, for those who liked the other two episodes, responses included, “because I love animals” for Episode 1 and “because I found this problem related to me” for Episode 2. Although there were also more simplistic responses (e.g., “because it was interesting” or “it was the most fun”), many of the responses related to a feeling of connection to that episode and an understanding for the characters’ feelings and situation. This ability to relate to the content may promote engagement with the game.

### Decision-making

Learning how to make better decisions and understanding the consequences of being a decision-maker emerged as a prominent theme (documented across 32.9% of student interviews). Participants reported that playing Quandary had improved their real-life decision making abilities and that they had gained new insight into what making decisions that would affect many people was like; for example, one girl said that “Yes, I think that … exactly, like I feel more independent now, like yeah I can make my own decision I don’t have to ask people because I can make my own conclusions about stuff.” Another child said that “If like problems come in my way throughout my life. I like kind of know how to solve them now.”

Decision-making was also described as both a highly enjoyable and highly challenging part of Quandary. For example, one child said that he enjoyed Quandary because “it’s like you can decide what you’re gonna do. You can decide which path you’re gonna take and which solution you’re gonna do.” Another participant enjoyed the complexity of the decision-making process, saying that “I liked the way that you have to solve the problems. There was a lot of fights that you had to like help. There was a lot of different opinions. So I liked that the best.” Making decisions that would please all the game’s characters was described as difficult, however: one child said that “Well, I thought it was pretty hard to satisfy everybody—to make a decision that everyone will be okay with.”

### Leadership

Many of our interviewees included the feeling of being a leader as an especially enjoyable aspect of Quandary (documented across 10.1% of participant interviews). Participants described how as the captain of the colony, their decisions were of great importance: “I liked how I was like, kind of the captain, sort of, and got to make the final – the most impact in the decision.” Another child similarly said, “I liked that every solution that there was I made the decision to make it, and like it’s what I decided to do that actually happened um that’s it. I felt like I was in charge.” Participants also discussed the feeling of having responsibility for the well-being of the colony: “I feel like the president or something like that because you have to make decision about it will be benefit the whole country or at least the part that they were living. And it was kind of fun. For a moment it feels like you are someone important.” Another said participant that said, “Well, it made me feel like I was a captain and it made me feel like I had, um, like I could take care of people.”

### Learning from Quandary

In our one-on-one interviews, we asked participants what, if anything, they felt they learned from Quandary. Several themes emerged from these data. Ninety-two percent of interviews contained at least one of the subsequent themes. These themes were unique to the students’ experience of playing Quandary; the following codes did not emerge in the control (Mayan Mystery) condition.

### Game-related components

The following describe the ways in which participants in the Quandary condition discussed components (and core objectives) of the game.
Morality-related components

We describe the themes that emerged related to issues of morality in the following section; that is, how did the players of Quandary describe

Importance of community. Another theme that emerged centered on the benefits of working with and for the entire community and the enjoyment of making a positive change in the virtual community. One child extensively discussed these ideas: “I liked that you have to do with a community. Like you make a decision working like with the agreement of the whole, entire community and try to make it better. I like, it’s kind of helpful for you to do in the future. For if you need to make any decision, you can ask other people. They will help you with it. It gives a lot of idea of what’s gonna happen if you do it in a team, which is – which I like.” The enjoyment of seeing the virtual colony succeed was expressed by another child: “It’s fun, and I like how it’s a mystery at the beginning, and then after—at the end—it’s like you get to, you did something really good for your community.” Another participant said that playing Quandary taught him about working together as a community: “That like people could disagree and agree with each other, but there’s always a solution.”

Consideration of others. A common theme across all of our interviews was the importance of considering the opinions and feelings of others when making decisions (documented across 21.5% of student interviews). This theme had two key aspects: first, that one’s decisions would be better if one took others’ ideas into account. For example, one child said that “if you had to decide it, without knowing what other people thought of it, you might make the wrong decision without doing it on purpose.” Another child suggested, “the decision would be better because not one person decides it, but a lot of people decide it.”

The second aspect of this theme was making decisions that ensured the well-being of everyone affected. Participants said that “everyone’s decision mattered and that like my decision is based on what everyone else said. ‘Cause I basically, like, I chose my solutions by like how many people agreed with each one so yea,” and, “I learned that picking one side isn’t fair to other people and that you shouldn’t just go with one person.”

Perspective taking. The final core theme that emerged from our interviews involved learning about perspective taking (found across 25.3% of participant interviews). Participants discussed the importance of hearing ‘both sides of the story’ and understanding the feelings of everyone involved many times, as in these quotes: “I, probably, learned that like everybody has a different opinion and not everybody like agrees with the same thing,” “Um, well, I liked how you could hear both sides of the story and I liked that because a lot of the times, when I get into arguments, um, like we don’t even bother to hear both sides of the story and so I found that really—and so I found it kind of interesting just to hear what all the different people had to say,” and “Um, I think I learned that you should always listen to both sides of the story because you’ll never know what other people have to say unless you listen to what they have to say.”

Qualitative depth coding

Analyses of the empathy related questions involved coding the depth of response to the questions. As previously noted, we used a global coding rating based on depth of response and compared levels of depth across Quandary and Control conditions.

Across schools, we found that students in the Quandary and Quandary + Facilitation condition showed greater depth of responses than students in the Control condition. The depth coding was conducted on all of the transcripts by trained graduate research assistants who were blind to the hypotheses. Reliability was satisfactory with a Cohen’s kappa coefficient = .83 (percentage agreement = 89%). It is important to note that, although the content of the student responses did not vary across conditions, the way in which students discussed issues of empathy and perspective taking was differentiated by condition.

Quantitative changes

As noted earlier (see the section The present study), we complemented our interviews with a quantitative analysis of participants’ questionnaire responses. As an initial step, we verified that the three groups (Quandary, Quandary + Facilitation, Control) did not significantly differ on any measure at pretest. Our analysis of the pre-test data revealed no significant differences across groups for any outcome (PYD: $\Lambda = .95$, $F(12, 310) = 0.69$, $p = .76$; AEC: $\Lambda = .96$, $F(6, 316) = 1.17$, $p = .32$; Prosocial: $\Lambda = .98$, $F(8, 304) = 0.39$, $p = .92$; Moral Reasoning: $\Lambda = .96$, $F(8, 304) = 0.69$, $p = .70$; Cyberball: $F(2,47) = 1.08$, $p = .35$).

To examine longitudinal changes, we next ran doubly multivariate MANOVAs. In these models we were specifically interested in the three-way interactions between multivariate responses, condition, and time, which indicated whether at least one outcome variable displayed significant cross-group differences in change between pre-test and post-test. Contrary to our hypotheses, none of these models suggested between-group
differences in change (PYD: $\Lambda = .94$, $F(12, 242) = 0.62$, $p = .82$; AEC: $\Lambda = .97$, $F(6, 250) = 0.66$, $p = .68$; Prosocial: $\Lambda = .96$, $F(8, 242) = 0.56$, $p = .81$; Moral Reasoning: $\Lambda = .98$, $F(8, 246) = 0.29$, $p = .97$; Cyberball (pre-post): $\Lambda = .99$, $F(1, 27) = 0.33$, $p = .97$; Cyberball (posttest only): $F(2,109) = 0.41$, $p = .67$). Finally, because prior research with adolescent responses to video games have reported gender differences (see Lucas & Sherry, 2004), we assessed whether such differences existed across these variables. No significant differences were found.

Discussion

Video games have been successfully used to teach cognitive and task-specific skills (e.g., D. R. Anderson & Kirkorian, 2015; Hattie, Biggs, & Purdie, 1996), but their potential as contexts for moral learning has not been extensively explored (Bers, 2012; Bers & Kazakoff, 2013; de Freitas & Griffiths, 2008; Mitchell & Savill-Smith, 2004). We examined how youth engaged with and were affected by Quandary, a moral learning game that challenges players to investigate complex problems with no clear solutions and make decisions that benefit the community.

Our first research objective was to understand if and how youth engaged with a moral learning game. The vast majority of our participants said that they enjoyed Quandary and would play it again in their free time, showing a high level of engagement with the game. This finding shows that a game focused on moral learning can engage youth and be an enjoyable play experience. We also found that participants’ enjoyment of the game was affected by the game content. When asked which episode of Quandary was their favorite, the majority of participants selected Episode 3, Fashion Faction. Participants’ reasons for choosing this episode as their favorite suggested that it resonated with them. At the time of data collection, two of the schools in this study were debating whether to implement a uniform policy. This situation, Fashion Faction’s focus on colony uniforms and dress codes may have been especially relevant to our participants’ lives.

Engaging and entertaining players is necessary, but not sufficient, for a game to be a successful learning tool. The game must also successfully communicate its ideas. Our qualitative interview and short answer data showed that when asked what they enjoyed about and learned from Quandary, youth responded with themes that captured Quandary’s intended messages. Participants reported the struggle related to the game, such as making decisions that would not please everyone, and being in a leadership position, as well as more explicit themes related to moral thinking, such as supporting one’s community, and caring for others, and taking the perspectives of others. As Damon (2008) noted, moral growth involves becoming aware of others’ needs, weighing your role in addressing those needs, and acting on it through care and support. It is notable that these data came from responses to open-ended questions about what they learned directly from the game; the fact that participants brought up these themes spontaneously shows that they were able to understand Quandary’s content and messages. In addition, participants reported that these aspects of the game were some of the things they enjoyed most about Quandary, showing that our participants were able to engage with complex moral concepts in an enjoyable way.

Overall, our findings suggest that a video game focused on moral learning can be enjoyable and engaging for youth. Further, such a game can offer opportunities to engage with complex ideas in a playful and entertaining way, yet challenging for how to be fair and just when there is not clear solution for individuals or the community. Participants in our study reported a positive experience in wrestling with the complex scenarios while making connections to their own experiences.

It should be noted that these themes were found across all participants in the Quandary conditions, without difference between those who had an additional facilitation discussion. This finding begets future research to understand why, as hypothesized, reflection after game sessions did not seem to elicit greater depth of moral themes. However, this finding gives preliminary evidence that the game affected their open-ended responses irrespective of the facilitation.

Despite these positive qualitative results, our quantitative analyses showed no changes in participants’ perspective taking, empathy, or helping behavior. There are several possible explanations for this discrepancy. The first is the amount of time playing the game. Although participants played Quandary for three weeks, each session was less than an hour long, leading to a very low dosage of game play. Previous video game training studies using action games have used larger dosages (involving 20, 30, and 50 hours of gameplay; e.g., Oei & Patterson, 2013). Furthermore, it is notable that our empathy scale was designed to assess trait empathy, a construct that may be unlikely to show changes over such a brief period of time. Our participants may not have spent enough time playing Quandary for changes to occur.

Finally, Quandary might be more effective if it were more explicitly connected to participants’ lives. As previously discussed, our participants were most engaged and reported the most enjoyment when playing episodes of Quandary that had personal relevance to them. In our interviews, we heard some participants...
making connections between the tasks of the game and solving problems in their own lives. Our depth-of-response coding of the interview data also showed that participants who participated in discussions after playing Quandary responded in more complex ways. Thus it seems possible that Quandary would be more effective if connections to players’ lives and ways to apply the skills of the game to real-life problems were made explicit by embedding Quandary in a larger curriculum. Future research should explore this possibility.

This study is limited in several ways, including the small sample, limited exposure to the game, varied contexts for game play, and the fact that we studied only middle school students from the northeastern U.S.; thus, findings with these youth might not generalize to youth living in other locations. The small sample did not allow us to run clustering analyses; therefore, we could not take into account the possible effects of nesting within classrooms and schools. Other limitations involve those of measurement. First, the competence component of the PYD measure was below a typically acceptable level. Although Cronbach’s alpha is a lower-bound estimate of reliability, and previous work has noted convergent and divergent validity of this measure despite low alpha levels (Geldhof et al., 2014), future research should consider alternative measures of this component of PYD. In addition, the selected measures were, perhaps, not sensitive to change and that we used an untested game to assess moral reasoning. Other methods assessing these constructs may provide different findings. For instance, the use of Kohlbergian-like dilemmas might be a fruitful method to employ (e.g., Colby et al., 1983). In turn, asking youth to identify moral exemplars and compare their own behavior to them has been shown to be a potentially useful approach to understanding and enhancing the moral development of youth (e.g., Johnson et al., 2016). Nevertheless, we believe that the findings from this research provide support for the ways in which Quandary promotes moral thinking and, more broadly, contribute to the study of and integration of digital games on learning, behavior, and educational contexts. This work contributes to the literature and efforts toward the promotion of video games as powerful and dynamic learning and educational tools.

In sum, video games have become part of the ecology of contemporary youth (Anderson & Kirkorian, 2015; Bers & Kazakoff, 2013). In light of the growing interest of educators to promote moral or character development (Berkowitz, 2012), incorporating ecologically valid experiences into their curricula would enhance the salience of and interest in such education (Lerner & Callina, 2014). We have focused on one moral development video game, Quandary, in the present research, and believe the present results encourage further research involving this game and, as well, the development and use of other such games in efforts aimed at enhancing moral learning and development.

References


